



Cornell University  
Cornell Institute for Social and Economic Research

# R<sup>2</sup>

Towards Fully Replicable Data Analysis in an Increasingly Connected World

Florio O. Arguillas, Jr., Ph.D.

William C. Block, Ph.D.

Jeremy Williams

CISER

2016 IASSIST Conference

June 1-3

Scandic Hotel

Bergen, Norway



International Association for Social Science  
Information Services & Technology

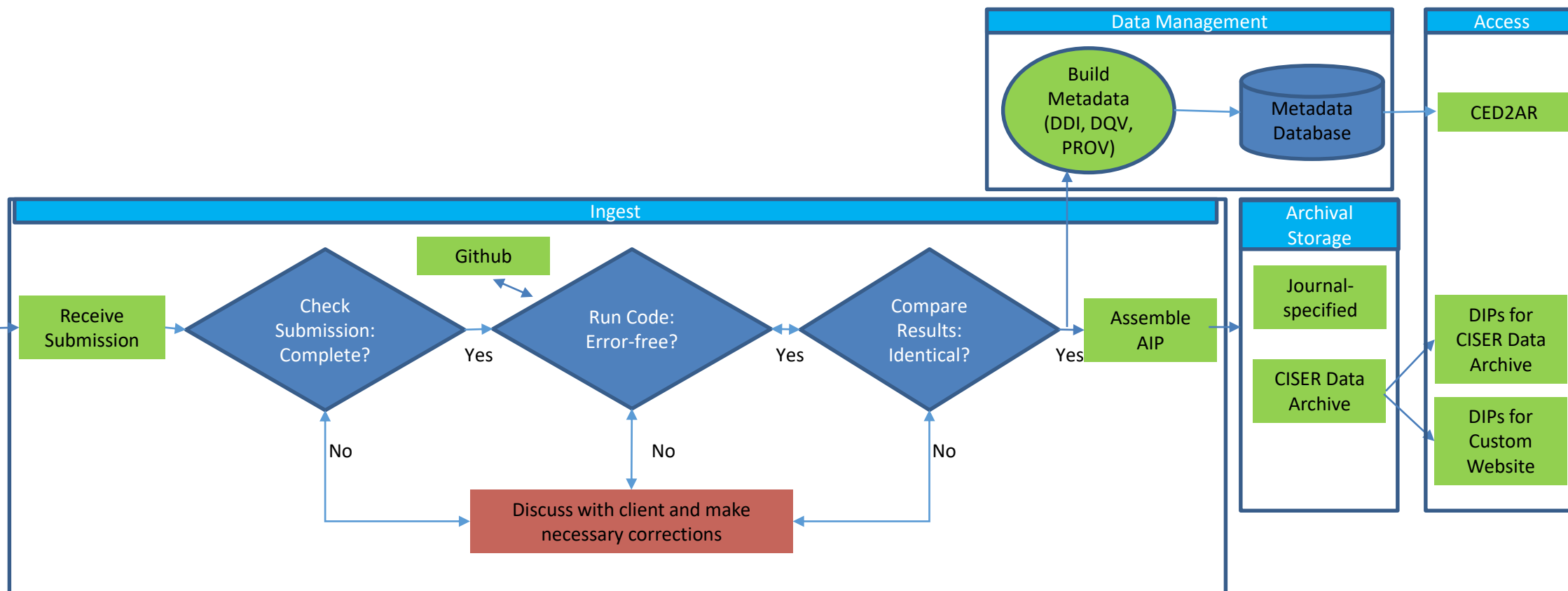


## Step 1

You've  
got Mail!

## Step 2

Client  
Deposits





	1990	1990	1990
1990	Wanted	Mistimed	Unwanted
Mistimed	Consistent	Inconsistent	Inconsistent
Unwanted	Inconsistent	Consistent	Inconsistent
	Consistent	Inconsistent	Consistent

FIG. 1—First Dependent Variable Construction

TABLE 1a				
Consistency of Women's Wantedness Reports Between Time 1 and Time 2 About the Same Pregnancy According to Pregnancy Outcome (Weighted Percentages)				
Pregnancy Outcome	1990-1995 Women Consistent with Report	1990-1995 Women Inconsistent	1990-1995 Women Consistent with Report	1990-1995 Women Inconsistent
Live birth	24.9	25.0	1338	1338
Abortion	28.3	71.7	140	140
Stillbirth, miscarriage, ectopic pregnancy	32.5	67.5	240	240
Total	26.3	73.7	1718	1718

TABLE 1b				
Consistency of Women's Wantedness Reports Between Time 1 and Time 2 About Partners' Views According to Pregnancy Outcome (Weighted Percentages)				
Pregnancy Outcome	1990-1995 Women Consistent with Report	1990-1995 Women Inconsistent	1990-1995 Women Consistent with Report	1990-1995 Women Inconsistent
Live birth	26.0	26.0	1250	1250
Abortion	30.0	69.9	120	120
Stillbirth, miscarriage, ectopic pregnancy	33.1	66.9	218	218
Total	28.0	72.0	1600	1600

rates for different Time 1 intention statuses. As a result, in those two tables the dependent variable can capture shifts toward more positive and more negative views. The dependent variable comprises three categories: More positive, more negative, and consistent. A more positive report resulted (1) if the woman claimed

at Time 1 that the pregnancy was unwanted, and then reported at Time 2 that the pregnancy was mistimed or intended, or (2) if she said at Time 1 that the pregnancy was mistimed and at Time 2 that it was intended. A more negative report resulted (1) if the respondent said at Time 1 that the pregnancy was intended and said

ion ▾ Processing and Analysis ▾ Command Files Search C

Name ^

Date modified

00\_masterdo 4/27/2016 3:34 PM

01\_run\_extract\_hhcomp 4/27/2016 2:30 PM

02\_recode\_hhcomp 4/27/2016 2:30 PM

03\_run\_extract 4/27/2016 2:30 PM

04\_recode

05\_run\_hier\_extract

06\_recode\_who

07\_recode\_activity

08\_recode

09\_asr\_r8r

```
26 /* Survey set the data to incorporate PSUs and strata */
27 svyset [pweight=perawgt], strata(sest) pau(secu)
28
29 /* Table 1: Descriptives—Overall sample */
30 /* 1995 vs. 2006-10 women descriptives (col's 1 and 2 in T1, descriptives) */
31 bys survey:el duration ucon2 ucon3 cohbirth black hispan other daded3 daded4 daded1 daded4 ///
32 noded3 noded4 noded5 noded6 instar church ///
33 lth high somecoll coll agemom ageday planned momprevkide priormar priorcoh anyprioruni ///
34 mistimed unwanted ageoldest1 ageoldest2 ageoldest3 babdoby priorunikids timebaby anyshotgun ///
35 if b10=1 [aweight=perawgt]
36
37
38 /* Table 2: Logistic regressions */
39 /* Model 1 */
40 local altvmar "newvmar"
41 foreach j in 1995 2007 {
42   svy: logit tvsplit tvduration cohunion cohbirth "altvmar" ///
43   if b10=1 & survey=="j", or
44 }
45 /* Test coefficients in Table 3 */
46 /* Row 1 */
```

Variables Manager

Filter variables here

Drag a column header here to group by that column.

#	Name	Label	Type	Format	Value label	Notes
1	rectype	Record Type	byte	%8.0g		
2	caseid	ATUS Case ID	double	%14.0f		
3	year	Survey year	long	%12.0g		
4	pernum	Person number (general)	byte	%8.0g	pernum_lbl	
5	lineno	Person line number	int	%21.0g	lineno_lbl	
6	month	Month of ATUS interview	int	%21.0g	month_lbl	
7	day	ATUS interview day of the week	byte	%21.0g	day_lbl	
8	wt06	Person weight, 2006 method...	double	%17.0f		
9	age	Age	int	%8.0g		
10	sex	Sex	byte	%21.0g	sex_lbl	
11	race	Race	int	%42.0g	race_lbl	
12	marst	Marital status	byte	%24.0g	marst_lbl	
13	educ	Highest level of school comple...	int	%47.0g	educ_lbl	
14	empstat	Labor force status	byte	%22.0g	empstat_lbl	
15	spousepres	Spouse or unmarried partner i...	byte	%38.0g	spousepres_lbl	
16	wb_resp	Well-Being Module Respondent	byte	%22.0g	wb_resp_lbl	
17	wbwt	Well-being Module final statist...	float	%15.6f	wbwt_lbl	

Variable properties

Name: rectype

Label: Record Type

Type: byte

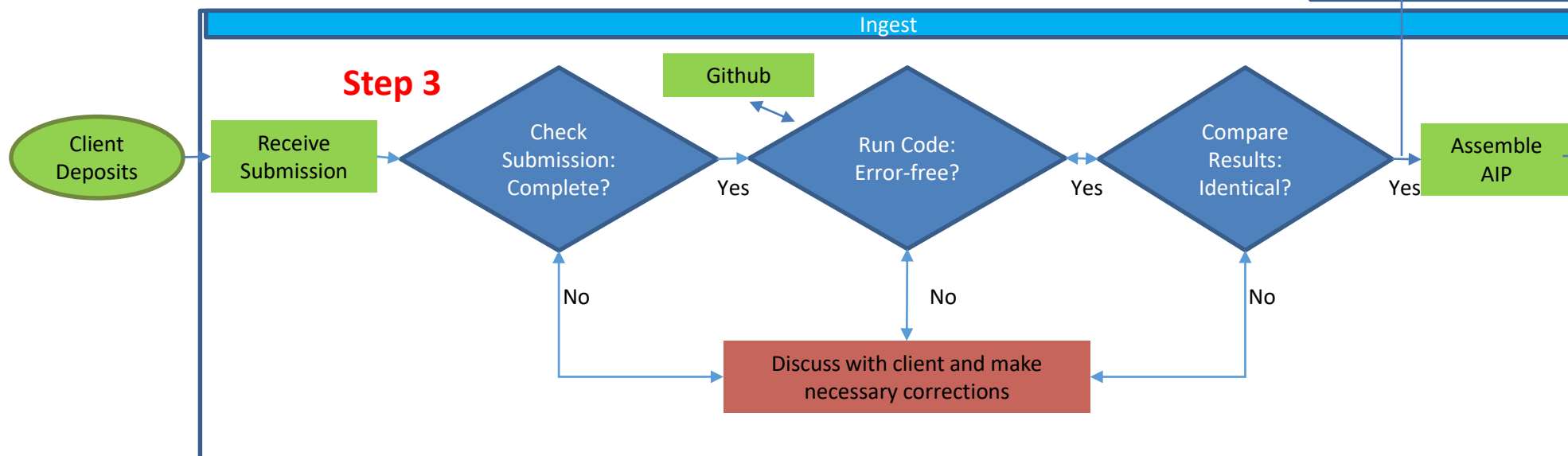
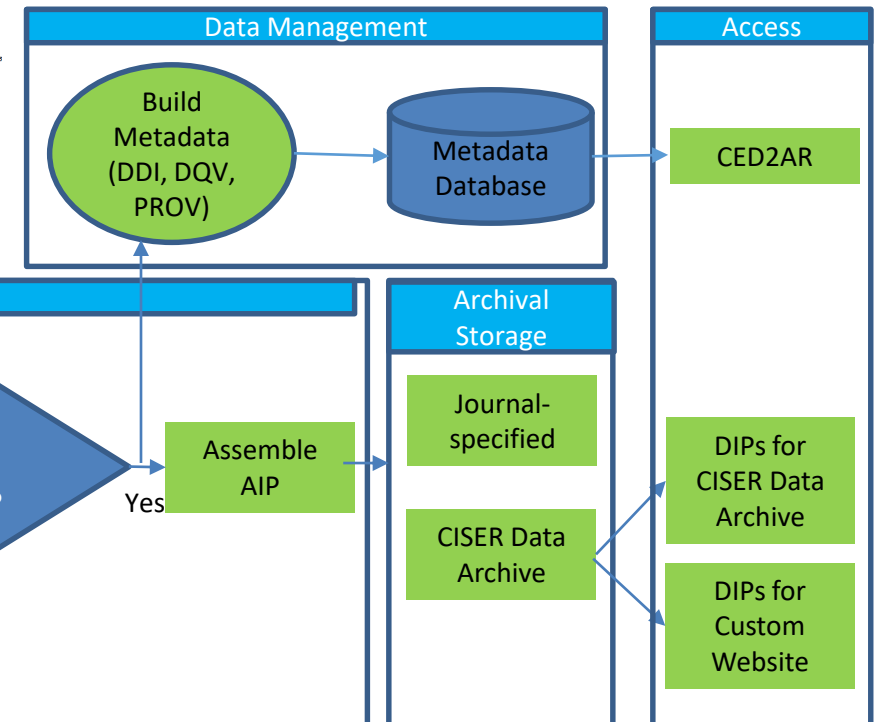
Format: %8.0g

Value label: Create...

Notes: No notes

Manage...

Vars: 399 CAP. NUM



COMPUTE TOTALINCOME = INCOME1 + INCOME2 + INCOME3.

COMPUTE TOTALINCOME = SUM(INCOME1, INCOME2, INCOME3).

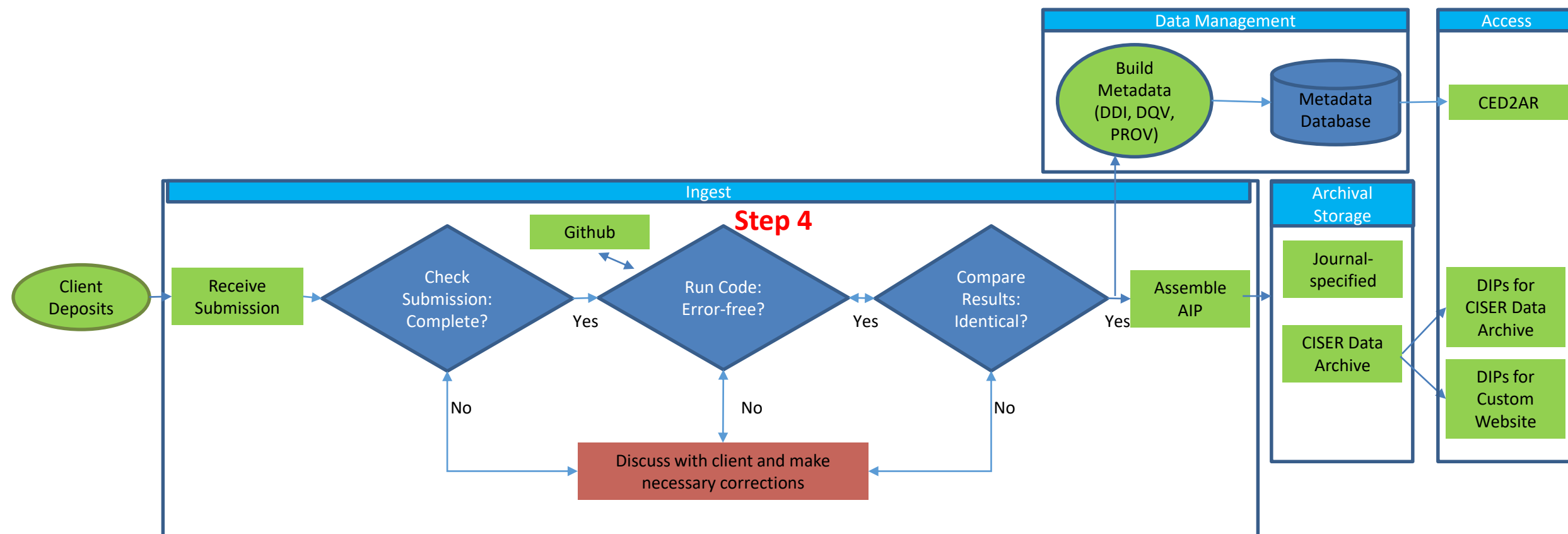




Table 1. Socio-demographic characteristics of parents (N individuals = 12,163)

	Full Sample	Mothers	Fathers	
N (individuals)	12163	7232	4931	actP
Age	38.38 (9.21)	36.91 (9.26)	40.22 (8.76)	age z zagesd
Number of children in the HH (reference = 1 child)				nkidhh31
2 children	39.50	39.18	39.91	nkidhh32
3+ children	20.44	18.85	22.43	nkidhh33
Age of youngest child (reference = <6)				ychild1
6-12	47.53	47.68	47.33	ychild2
13-17	32.83	31.71	34.22	ychild3
Race/ethnicity (reference = White, non-Hispanic)				hr1
Black, non-Hispanic	9.88	11.72	7.60	hr2
Hispanic	20.11	20.43	19.71	hr3
Other, non-Hispanic	6.57	5.80	7.53	hr4
Spouse or partner in the HH	82.80	74.10	93.64	pinhh
Household income (reference = <\$25,000)				inccat1
\$25,000-\$99,999	18.00	21.32	13.87	inccat2
\$100,000 +	57.70	56.32	59.42	inccat3
Income missing	23.14	20.78	26.08	inccat4
Employment status (reference = full time)				ws1
Part-time work	62.43	44.69	84.51	ws2
No paid work	14.32	21.28	5.67	ws3
College degree +	35.01	34.86	35.19	collgrad
Enrolled in school	5.82	7.77	3.39	mschool

Notes: 2010, 2012, and 2013 ATUS well-being sample, men and women with children in the household. N's are unweighted; means/percentages are weighted (using individual weights for individual-level and activity weights for activity-level). Standard deviations in parentheses. HH = household.

\* = Mothers' characteristics are significantly different (p<.05) from fathers'. (+test 'v')

```
.. **print Table 1
.. disp _varname|fs|men|women|diff m v. w"
.. _varname|fs|men|women|diff m v. w
.. disp _actP|'Nt'|'Nt1'|'Nt2'"
.. _actP|12163|4931|7232
..
..   foreach v in `alvars' {
..       disp "`v'|`v'F'|`v'1'|`v'2'|`test'v'"
..   }
age|38.4752||40.3477||36.9548||0
collgrad|.3478||.3479||.3478||.9938
hr1|.61640000000000000001|.64240000000000000001|.5954||.0001
hr2|.1033||.0772||.1245||0
hr3|.2131||.2117||.2142||.8134
hr4|.0672||.0688||.0659||.6115
inccat1|.1848||.1364||.224||0
inccat2|.56280000000000000001|.58620000000000000001|.54390000000000000001|.0002
inccat3|.2421||.2691||.2202||0
inccat4|.0102||.0082||.0119||.1349
inschool|.0568||.0338||.0756||0
nkidhh31|.4048||.3895||.4173||.0152
nkidhh32|.3886|.3929|.3851|.4867
nkidhh33|.2066||.2176||.1976||.0379
pinhh|.82610000000000000001|.9323||.7399||0
ws1|.6147||.82380000000000000001|.445||0
ws2|.1418||.0624||.2062||0
ws3|.2435||.1138||.3488||0
ychild1|.4681||.464||.4713||.52660000000000000001
ychild2|.3287||.3334||.3248||.425
ychild3|.2033||.2026||.2038||.9012
..
..   foreach v in age {
..       disp "zz'v'sd|(`v'sd')|(`v'sd1')|(`v'sd2'"
..   }
z zagesd|(9.08)||(.8.63000000000000000001)||(.9.14000000000000000001)
..
..   } //end t1
```

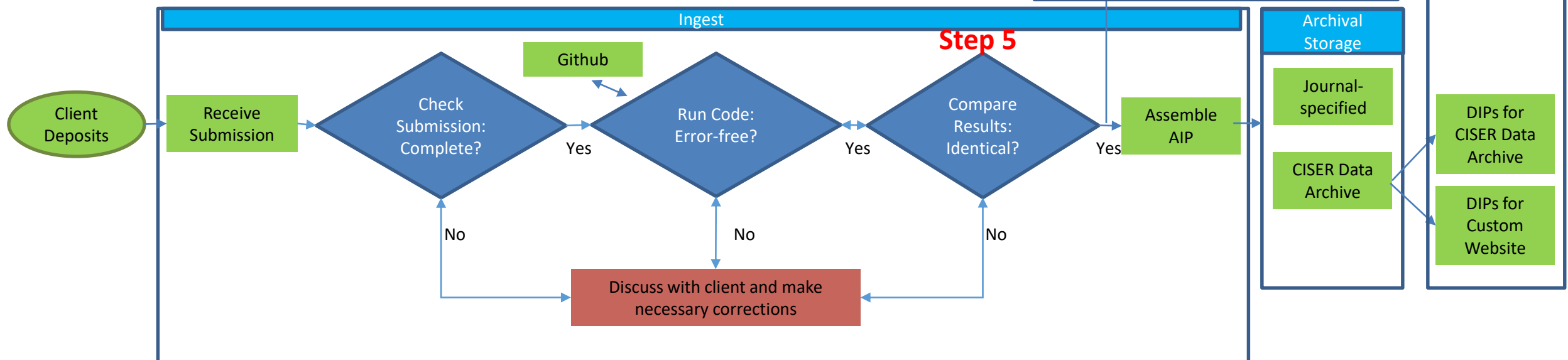






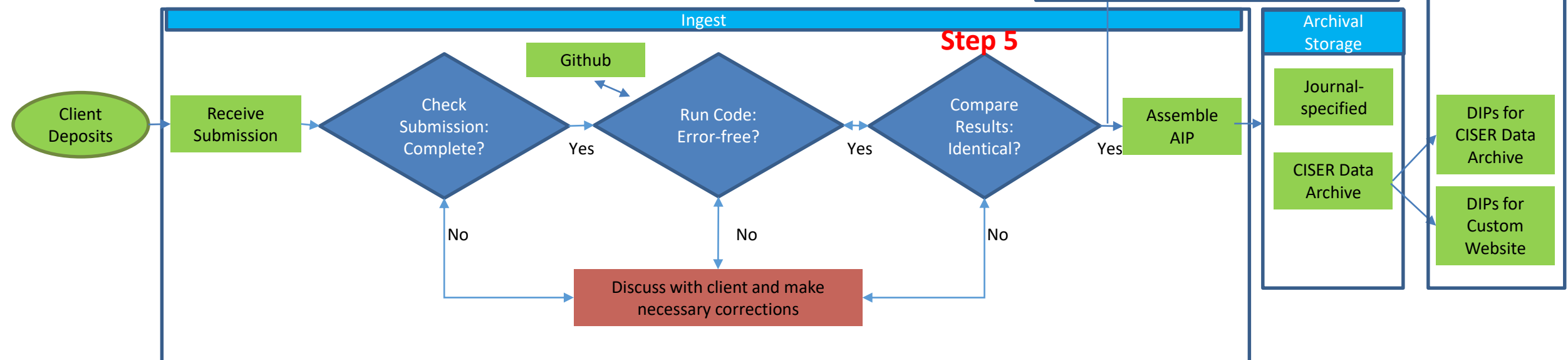
Table 1. Sociodemographic characteristics of parents (N = 12,163)

	Full Sample	Mothers	Fathers	Test v
N (individuals)	12163	7232	4931	-actP
Age	38.38	36.91	40.22	age
Number of children in the HH (reference = 1 child)	9.21	9.26	8.63	zzagesd
2 children	40.05	41.97	37.67	nkidhh32
3+ children	39.50	39.18	39.91	nkidhh33
Age of youngest child (reference = <6)	20.44	18.85	22.43	ychild1
6-12	47.53	47.68	47.33	ychild2
13-17	32.83	31.71	34.22	ychild3
Race/ethnicity (reference = White, non-Hispanic)				hr1
Black, non-Hispanic	63.44	59.54	65.16	hr2
Hispanic	9.88	11.72	7.60	hr3
Other, non-Hispanic	20.11	20.43	21.17	hr4
Spouse or partner in the HH	82.80	82.61	83.99	pinhh
Household income (reference = <\$25,000)				inccat1
\$25,000-\$99,999	18.00	21.32	13.87	inccat2
\$100,000 +	57.70	56.32	58.62	inccat3
Income missing	23.14	20.78	26.91	inccat4
Employment status (reference = full time)				ws1
Part-time work	62.43	44.69	84.51	ws2
No paid work	14.32	21.28	5.67	ws3
College degree + Enrolled in school	35.01	34.86	35.19	collgrad
	5.82	7.56	3.39	inschool

Notes: 2010, 2012, and 2013 ATUS well-being sample, men and women with children in the household. N's are unweighted; means/percentages are weighted (using individual weights for individual-level and activity weights for activity-level). Standard deviations in parentheses. HH = household.  
\* = Mothers' characteristics are significantly different (p<.05) from fathers'. (test v)

```
.. **print Table 1
.. disp _varname|fs|men|women|diff m v. w"
.. _varname|fs|men|women|diff m v. w
.. disp _actP|Nt|Nt1|Nt2|
.. _actP|12163|4931|7232|

1 2 3 4 5
foreach v in age {
  disp "zz'v'sd|('v'sd')|('v'sd1')|('v'sd2')|
  zzagesd| (9.08) | (8.630000000000001) | (9.140000000000001)
} //end t1
```



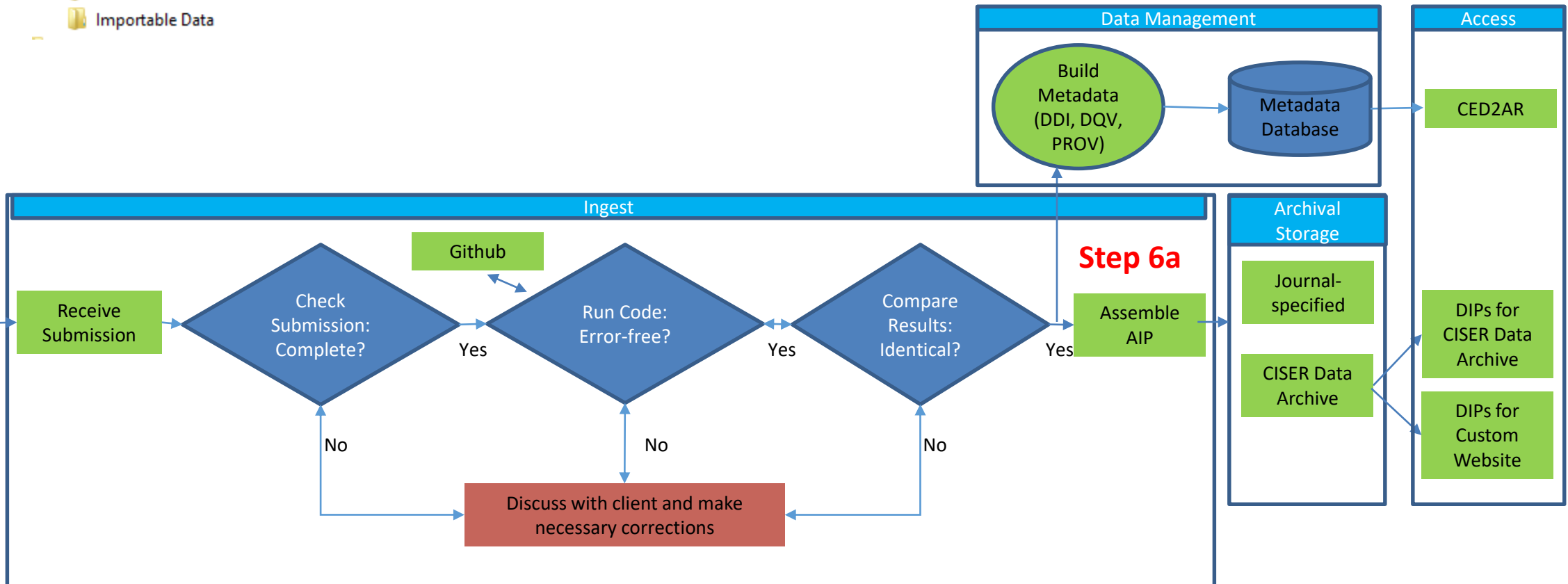
# Teaching Integrity in Empirical Research (TIER) protocol

<https://www.haverford.edu/project-tier/protocol-v2>



International Association for Social Science  
Information Services & Technology

- Replication Documentation
  - Original Data and Metadata
    - Metadata
    - Original Data
  - Processing and Analysis
    - Analysis Data
    - Command Files
    - Importable Data

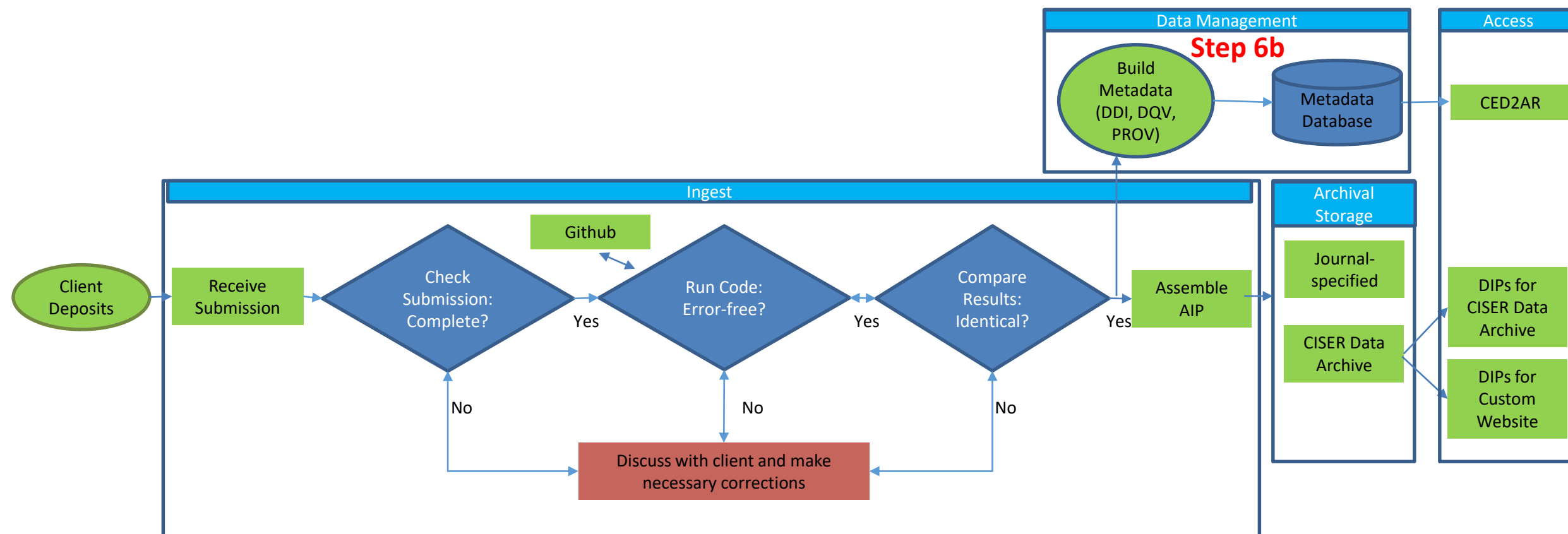




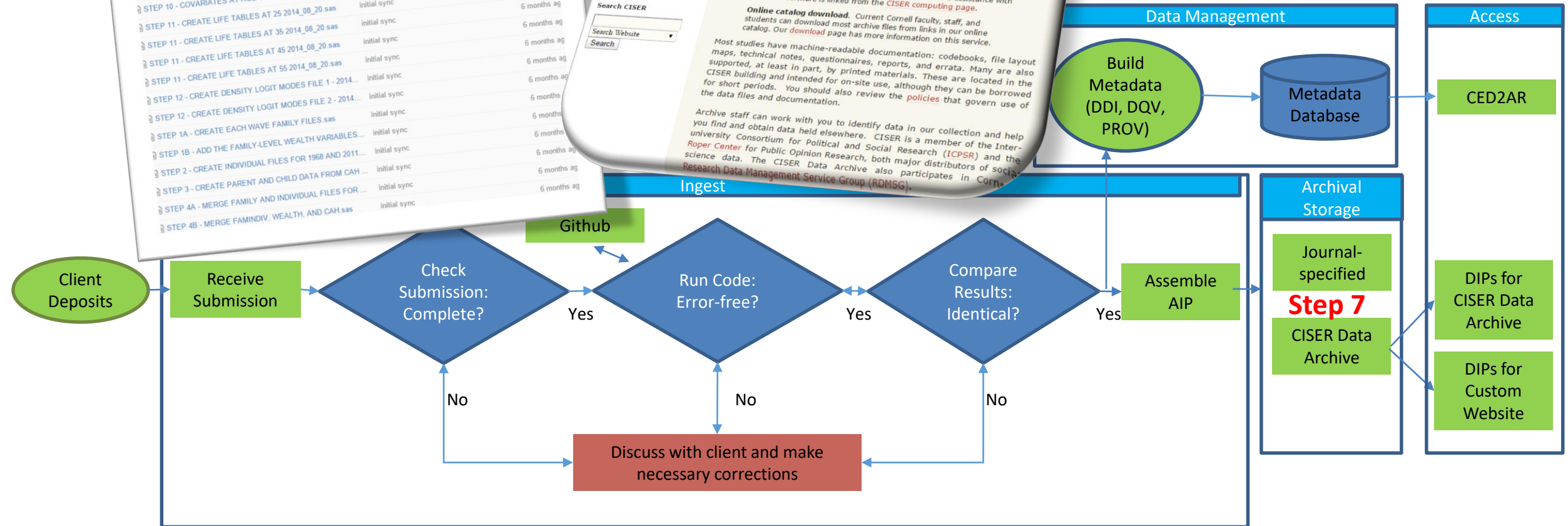
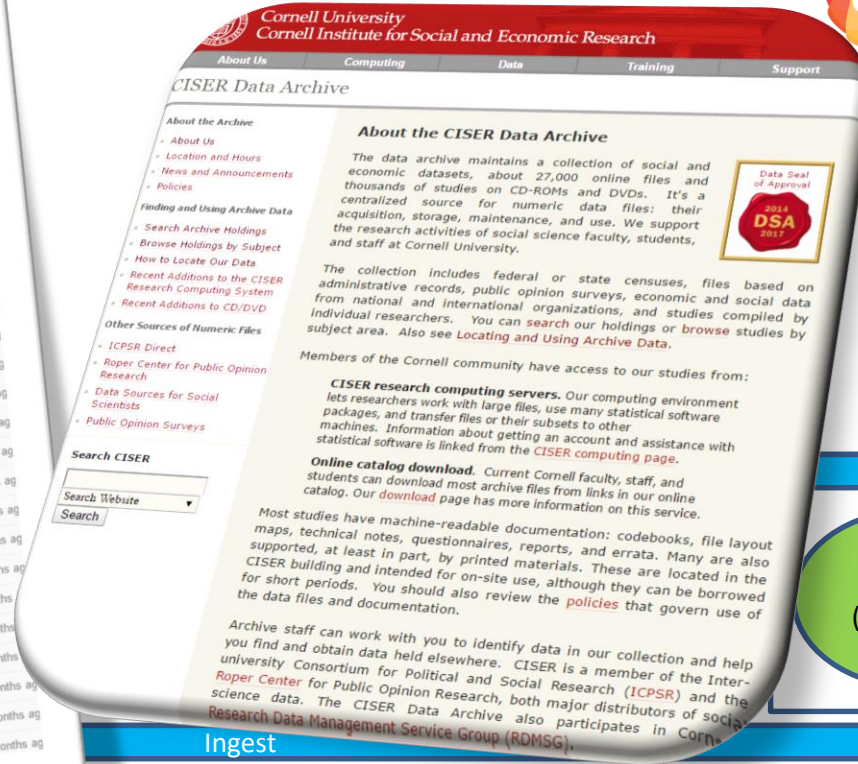
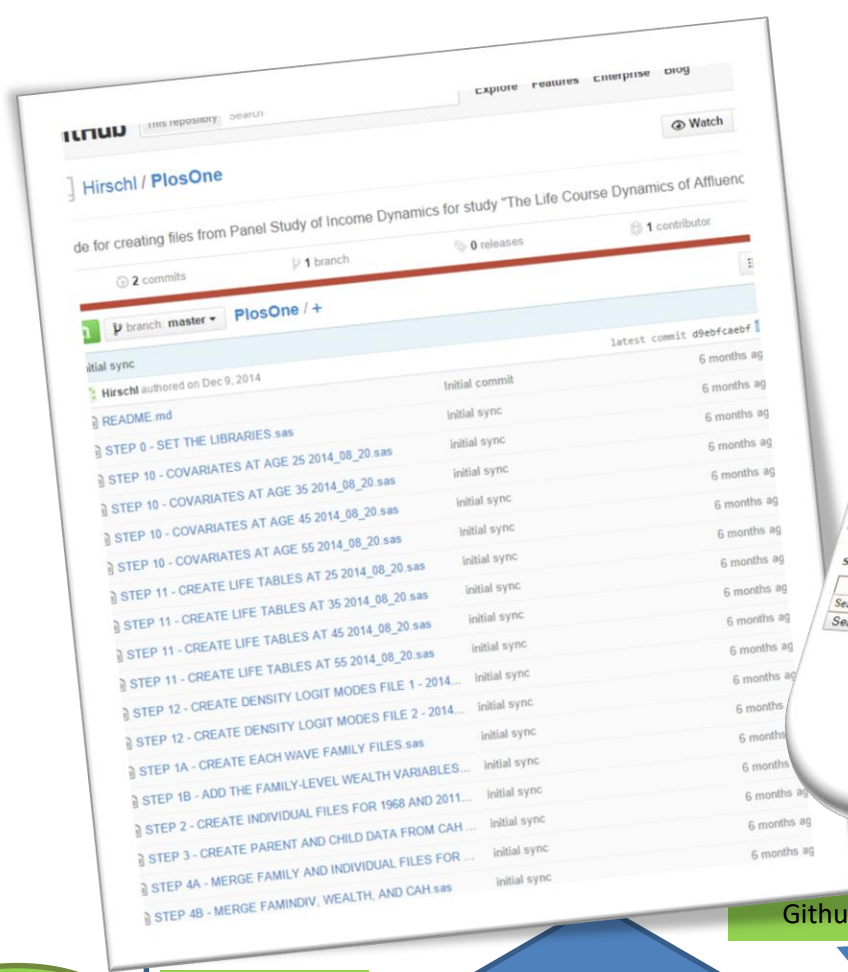
## Comprehensive Extensible Data Documentation and Access Repository (CED<sup>2</sup>AR)

Data Quality Vocabulary (DQV): <http://w3c.github.io/dwbp/vocab-dqg.html>

PROV: <http://www.w3.org/TR/prov-primer/>









Cornell University  
Cornell Institute for Social and Economic Research

About Us Computing Data Training Support

CISER Data Archive: Online Catalog

About the Archive

- About Us
- Location and Hours
- News and Announcements
- Policies

Finding and Using Archive Data

- Search Archive Holdings
- Browse Holdings by Subject
- How to Locate Our Data
- Recent Additions to the CISER Research Computing System
- Recent Additions to CD/DVD

Other Sources of Numeric Files

- ICPSR Direct
- Roper Center for Public Opinion Research
- Data Sources for Social Scientists
- Public Opinion Surveys

Search CISER

Search Website

Search

How parents fare: Mothers' and fathers' subjective well-being in time with children

Bibliographic Information:

Musick, Kelly, Ann Meier, and Sarah Flood (2016). Codebook: R2-MUSICK-002(2016). This study includes files created by Cornell researchers and/or staff.

View Abstract

Abstract: The shift to more time-intensive and child-centered parenting in the U.S. is widely assumed to contribute to child development, but implications for adult well-being are less clear. We assess multiple dimensions of parents' activities with children and explore how the gendered nature of time potentially contributes to differences in mothers' and fathers' parenting experiences. Relying on nationally representative time diary data linked to respondents' feelings in activities from the 2010, 2012, and 2013 well-being module of the American Time Use Survey (N = 12,163 persons and 36,036 activities), we find that parents consistently report greater subjective well-being in activities with children than without. Mothers, however, report less happiness, more stress, and greater fatigue in time with children than fathers. These gaps are relatively small and can be accounted for by differences in the activities that mothers and fathers engage in with children, whether other adults are present, and the quality of their sleep and leisure. We go beyond prior work on parental happiness and life satisfaction to document how contemporary parenting is woven differently into the lives of mothers and fathers.

Key words: parenting, subjective well-being, gendered family roles, time use

\* Direct correspondence to: Kelly Musick, 254 MVR, Cornell University, Ithaca, NY 14853; phone 607-255-6067; email [kmusick@cornell.edu](mailto:kmusick@cornell.edu). Earlier versions of this work were presented at the 2013 Annual Meeting of the Population Association of America, the Maryland Population Research Center, the 2015 Alpine Population Conference, the Centre for Economic Demography Seminar at Lund University, and the Swedish Institute for Social Research SWS Seminar at Stockholm University. We are thankful to the ASR editors and reviewers, as well as conference and seminar participants for many useful suggestions, in particular Francesco Billari, Rachel Dunifon, Karin Hallidén, Bo Malmberg, Jeff Neilson, Liana Sayer, Kammi Schmeer, and Maria Stanfors. We gratefully acknowledge seed grants from the Cornell Population Center and Cornell's Institute for Social Sciences and support from the Minnesota Population Center (5R24HD041023) and the American Time Use Survey Data Extract Builder project (5R01HD053654), funded through grants from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD).

go back new search browse archive holdings

Contact CISER  
Copyright © 1997-2019 Cornell Institute for Social and Economic Research  
Site Map

Voice: (807) 255-4801  
Fax: (807) 255-9353

Links to:

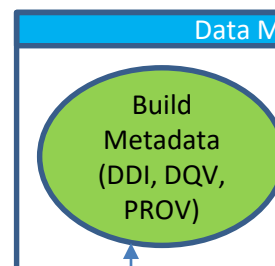
- Metadata with variable level search through CISER's instance of CED2AR
- Journal article
- Replication materials:  
V:\R2\<1stAuth>\<r2>

Social Science  
ology

- Replication Documentation
- Original Data and Metadata
- Metadata
- Original Data
- Processing and Analysis
- Analysis Data
- Command Files
- Importable Data

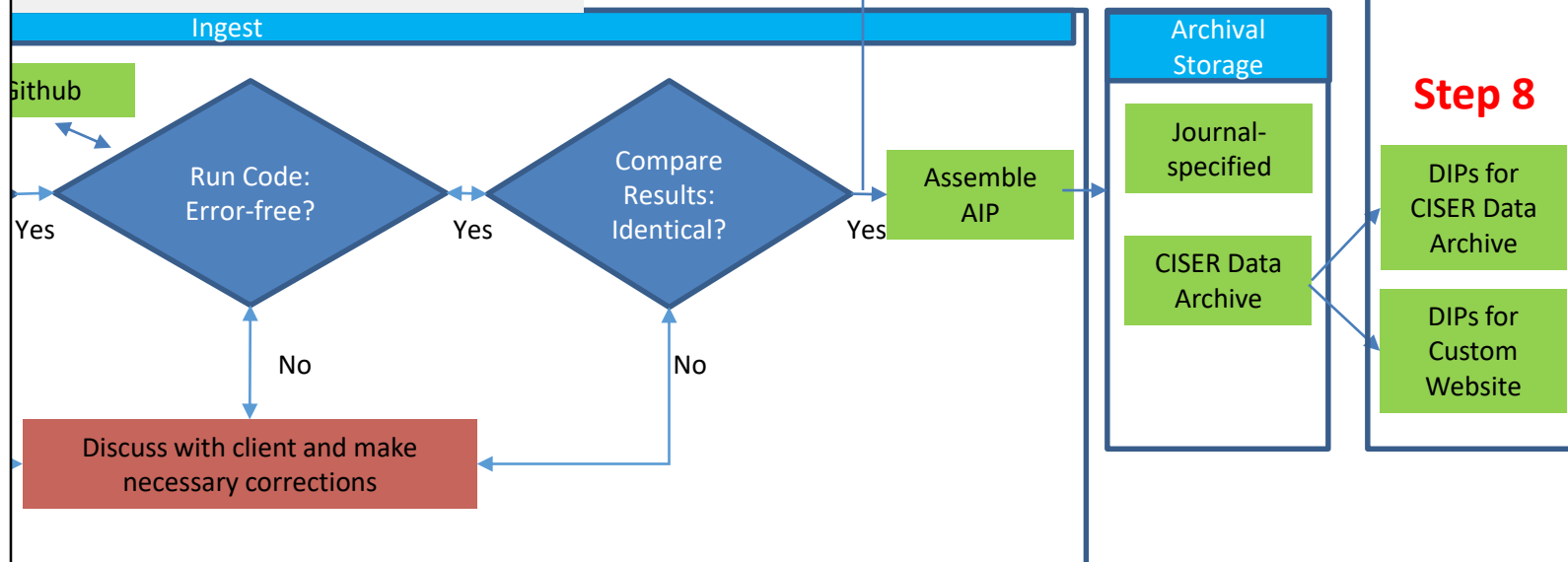
Access

CED2AR



Will also include:

- Persistent URL:  
<http://doi.org/10.6077/J5/CISER2775>
- Study citations
- Data citation
- Documentation about known quality issues
- User Feedback Form
- Forum or mailing list
- Social Media Accounts (e.g., Facebook, Twitter)





### About the Book

The United States has been epitomized as a land of opportunity, where hard work and skill can bring about personal success and economic well-being. Yet in their pursuit of the American Dream, many will experience poverty firsthand. Just how real that risk can be found in our new poverty risk calculator.

[Learn more »](#)

### Risk Calculator

Have you ever wondered what your risk of poverty might be in the future? Our calculator will estimate your personal risk over the next 5, 10, or 15 years.

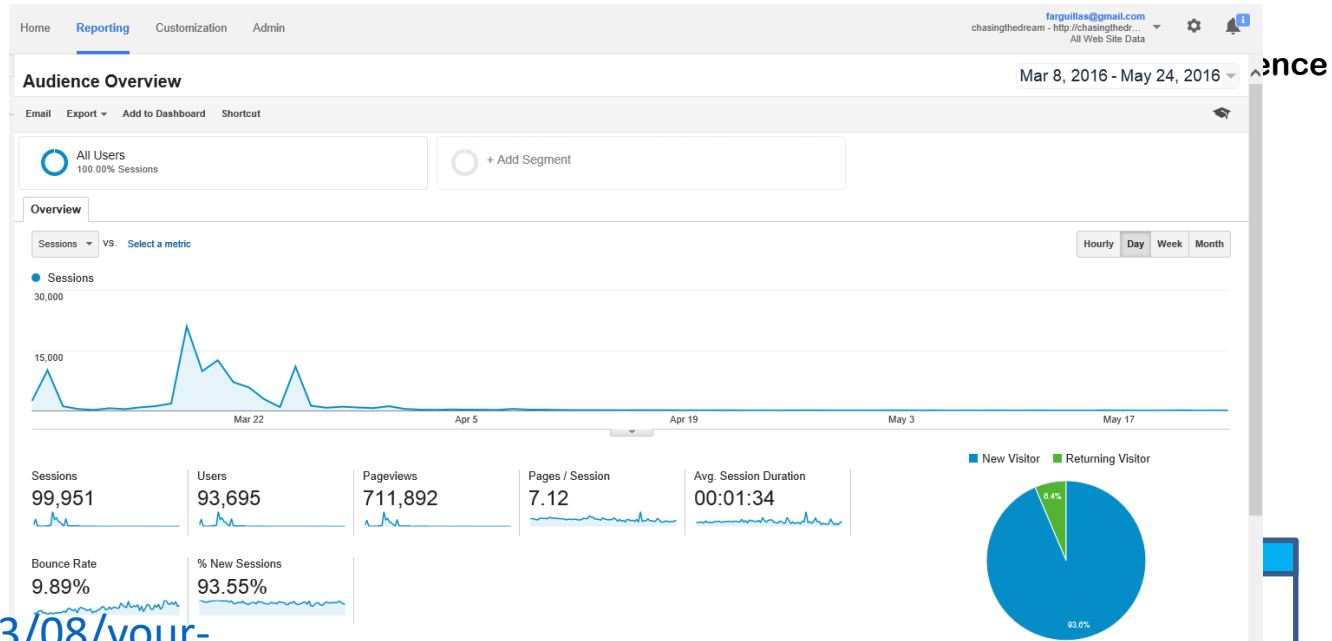
[Go to Calculator »](#)

### About the Authors

This work was a collaborative effort between:

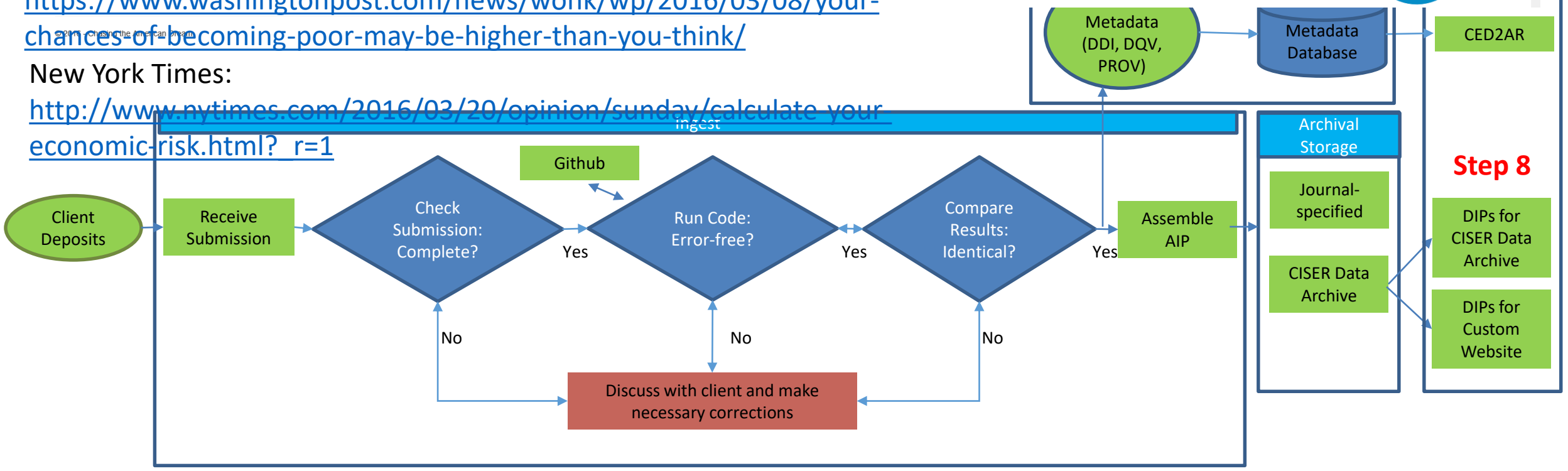
- Mark Robert Rank, Washington University
- Thomas A. Hirschl, Cornell University
- Kirk A. Foster, University of South Carolina

[Learn more »](#)



Washington Post:  
<https://www.washingtonpost.com/news/wonk/wp/2016/03/08/your-chances-of-becoming-poor-may-be-higher-than-you-think/>

New York Times:  
[http://www.nytimes.com/2016/03/20/opinion/sunday/calculate-your-economic-risk.html?\\_r=1](http://www.nytimes.com/2016/03/20/opinion/sunday/calculate-your-economic-risk.html?_r=1)





# Conclusion

- Process is time-intensive, costly, and confusing if data and code quality is low
- Suggest code review, use of versioning software, adherence to TIER protocol in managing files
- Reach out to Cornell graduate students and provide training on how to prepare research materials for reuse thru classes or workshops.
- CURATE AS YOU CODE AND CODE WITH REUSE IN MIND.
- For archives:
  - Incorporate DQV and PROV
  - Follow TOP Guidelines, Open Data Certificate Requirements, and TIER Protocol for managing and packaging replication materials
  - Anything else?



# Thank you!

- Contact author: Florio Arguillas (foa2@cornell.edu)

